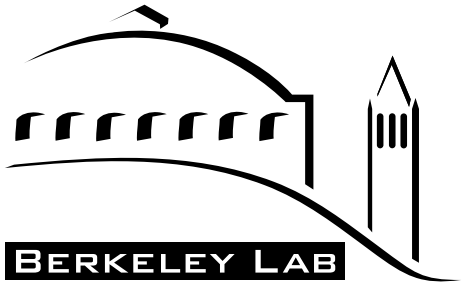
# High-Speed Distributed Data Handling for HENP: Architecture and Implementation[1]

*Brian Tierney[2], Craig Tull, William Johnston, and William Greiman*
*Information and Computing Sciences Division*
*Douglas Olson, Nuclear Science*
*Ernest Orlando Lawrence Berkeley National Laboratory*
*University of California*




1. This work is supported by the Director, Office of Energy Research, Office of Computation and Technology Research, Mathematical, Information, and Computational Sciences Division, of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098 with the University of California, and by ARPA ISTO

2. bltierney@lbl.gov, 510-486-5014, http://www-itg.lbl.gov/~johnston

**Goal is to demonstrate a scalable and flexible approach to the problem of high-bandwidth data handling for analysis of high-energy and nuclear physics data that will enable distributed computation and storage facilities to handle 20+ mbyte/s data streams (approximately the data rate needed for STAR analysis).**

**Objectives and accomplishments:**

- **The design and demonstration of a high-performance, application-level cache (DPSS) and tertiary storage system interface that will enable the analysis of all STAR summary tape data in "real-time" (at the data rate that summary tapes are produced)**
  - **the DPSS has supplied data to many parallel instances of the STAR analysis code at aggregate rates in excess of 20 megabytes/s**

- an interface between HPSS parallel tape drives and the DPSS has been designed
- a prototype operational system with these capabilities is being put into place in cooperation with the NERSC Mass Storage System group

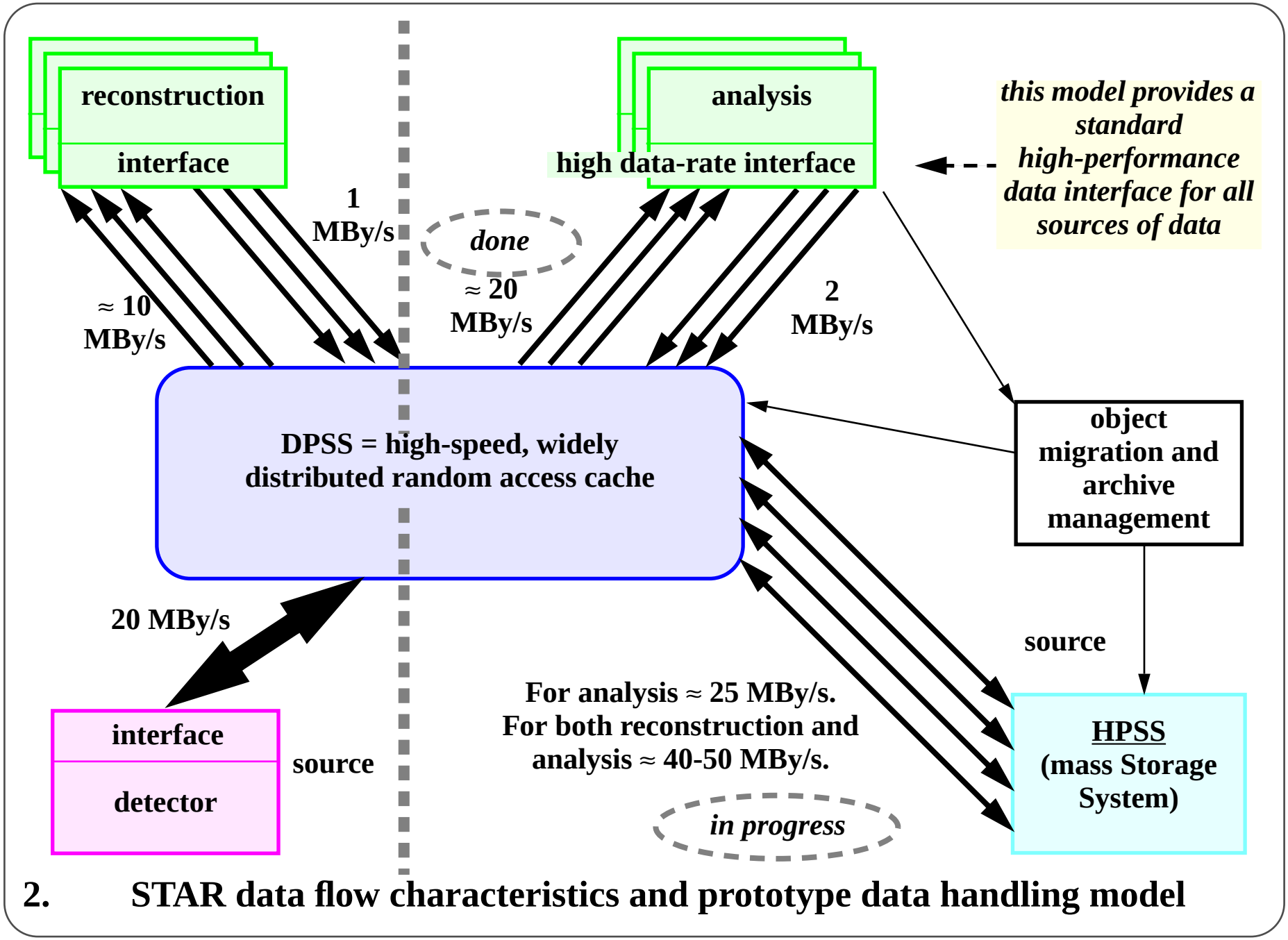
♦ A secondary objective is to demonstrate the potential of distributed storage and computational systems to do the first level of data analysis ("reconstruction") in real-time

- single 30 megabyte/s application-level data streams have been generated, routed through OC-12 (622 megabit/s) networks, and consumed by single or multiple processes on a Sun SMP

# The Overall Model

- The high-speed data handling model is based on the idea of a standard interface to a large, application-oriented, on-line cache.
- Each data source deposits its data in the cache, and each data consumer takes data from the cache, usually writing processed data back to the cache.
- A tertiary storage system manager migrates data to and from the cache.
- Depending on the size of the cache relative to the objects of interest, the storage system manager may move objects to the cache over some relatively long period of time; that is, the cache can serve as a moving window on the object/dataset.

# Model



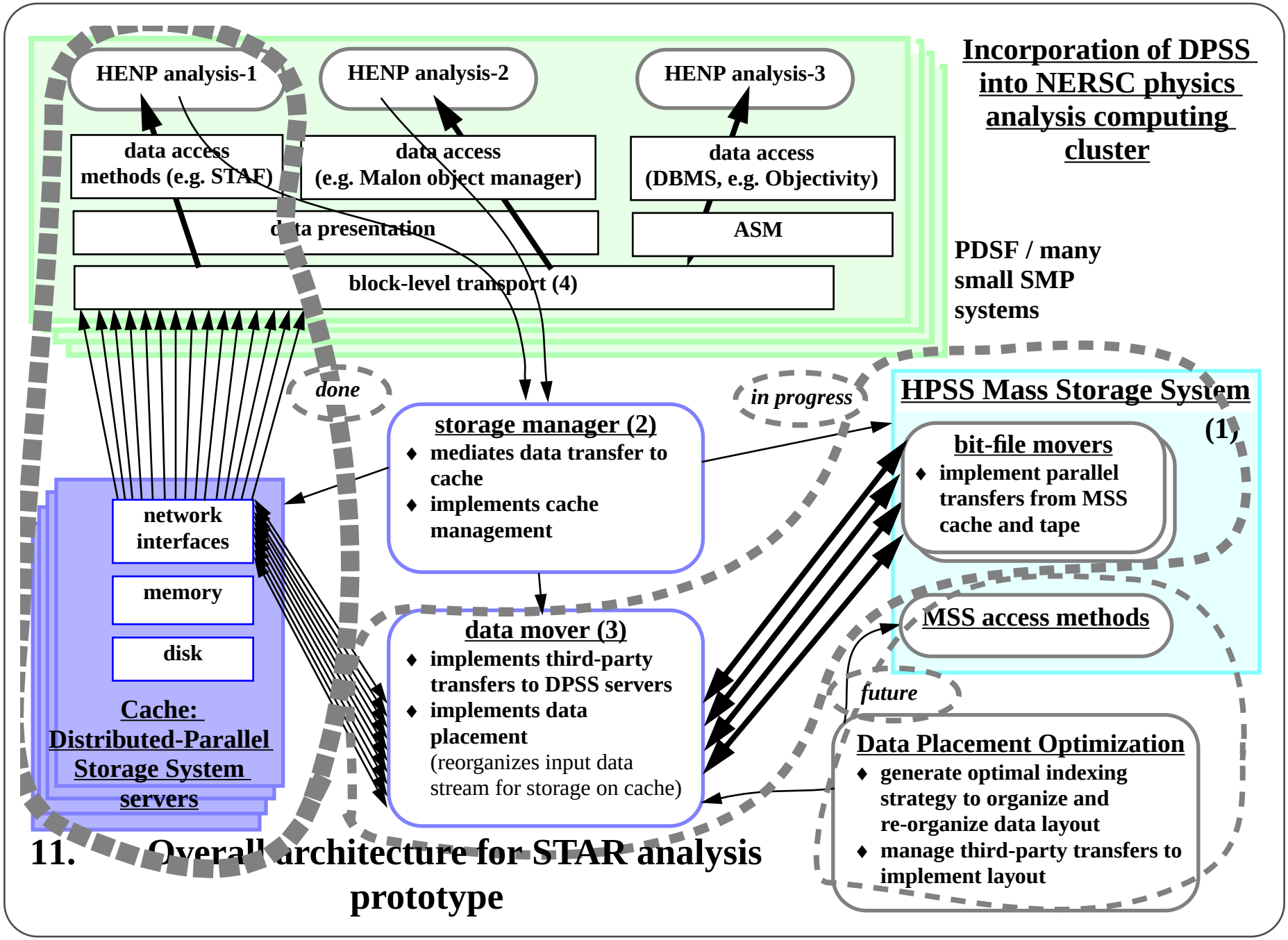

2. STAR data flow characteristics and prototype data handling model

# The Experimental STAR Analysis Configuration

## Overall architecture (figure 11.):

- The prototype analysis environment assumes that a high-speed tertiary storage system is the source of data (i.e. STAR reconstructed events) (1)
- Data is migrated from tertiary storage to the cache as a result of requests from the application (2,3)
- The cache management (2) mediates the movement of data from tertiary storage to the cache as it is consumed by the application
- The application uses a standard interface to access all data, whether in the cache, or waiting to be migrated in

# Prototype STAR Analysis Configuration

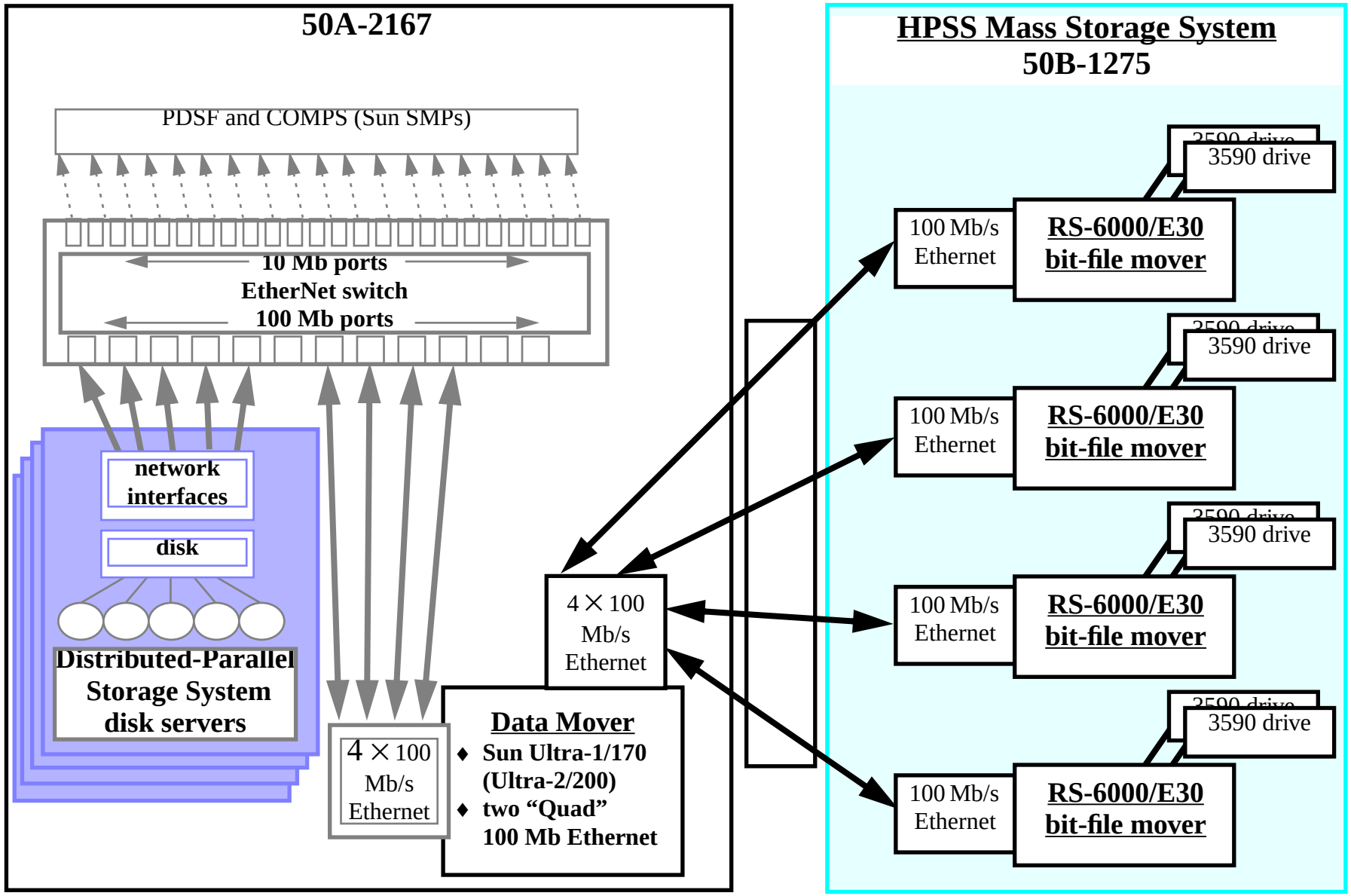



11. Overall architecture for STAR analysis prototype

# Prototype STAR Analysis Configuration

## Phase one implementation and experiment:

# Prototype STAR Analysis Configuration




**13. HPSS / DPSS experiment configuration - phase 1**

# Prototype STAR Analysis Configuration

## Phase two:

- Add STAF-based data filter on the Data Mover system (module (5) in

## See http://www-itg.lbl.gov/STAR